PATENT ABSTRACTS OF JAPAN

(11)Publication number :   06-149290

(43)Date of publication of application : 27.05.1994

---

(51)Int.Cl.   G10L   3/02

---

(21)Application number : 04-293412   (71)Applicant :   SANYO ELECTRIC CO LTD

(22)Date of filing :   30.10.1992   (72)Inventor :   TSURUFUJI SHINICHI

---

(54) SPEECH RECOGNIZING DEVICE

(57)Abstract:

PURPOSE: To improve the recognition rate of the speech recognizing device in noisy environment by absorbing the distortion of a feature parameter inputted to a microphone for speech input without being affected by the position of a noise source in a cabin and the position of the microphone for speech input, and efficiently performing noise removal. CONSTITUTION: The speech recognizing device is equipped with the microphone 1 to which a speech and an acoustic noise are inputted indirectly through a space, a sound signal input terminal 3 to which they are inputted directly from an acoustic device, an acoustic characteristic correcting means (sound field correction part 8) which compensates characteristics of the direct signals, a 1st feature extraction part 2 which extracts a 1st feature parameter by analyzing the input signal from the microphone 1, a 2nd feature extraction part 4 which extracts a 2nd feature parameter by analyzing the corrected input signal from the sound field correction part 8, a noise removal part 5 which generates a 3rd feature parameter by subtracting the 2nd feature parameter from the 1st feature parameter, a speech pattern generation part 6 which generates a speech pattern from the 3rd feature parameter, and a recognizing process part 7 which recognizes the speech pattern.

LEGAL STATUS

[Date of request for examination]   13.07.1999

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

CLAIMS

[Claim(s)]
[Claim 1] The voice recognition unit characterized by providing the following. The microphone into which the above-mentioned acoustic noise is indirectly inputted through the above-mentioned space while the voice uttered for speech recognition is inputted in the voice recognition unit which performs speech recognition in the space environment where the sound from sound equipment serves as noise for speech recognition. The acoustic signal input terminal into which the signal of the above-mentioned acoustic noise is directly inputted from sound equipment, without minding the above-mentioned space. It is a property amendment sound property amendment means about the direct signal of an acoustic noise inputted from the above-mentioned acoustic signal input terminal so that the property difference between the indirect signal of the acoustic noise obtained from the above-mentioned microphone and the direct signal of the acoustic noise obtained from the above-mentioned acoustic signal input terminal may be reduced. A 1st feature-extraction means to analyze the input signal from the above-mentioned microphone, and to extract the 1st feature parameter, A 2nd feature-extraction means to analyze the input signal rectified from the above-mentioned sound property amendment means, and to extract the 2nd feature parameter, A normal-mode-rejection means to deduct the 2nd feature parameter of the above from the 1st feature parameter of the above, and to generate the 3rd feature parameter, A voice pattern creation means to create a voice pattern from the 3rd feature parameter of the above, and a recognition processing means to perform recognition processing to the voice pattern created with the above-mentioned voice pattern creation means.
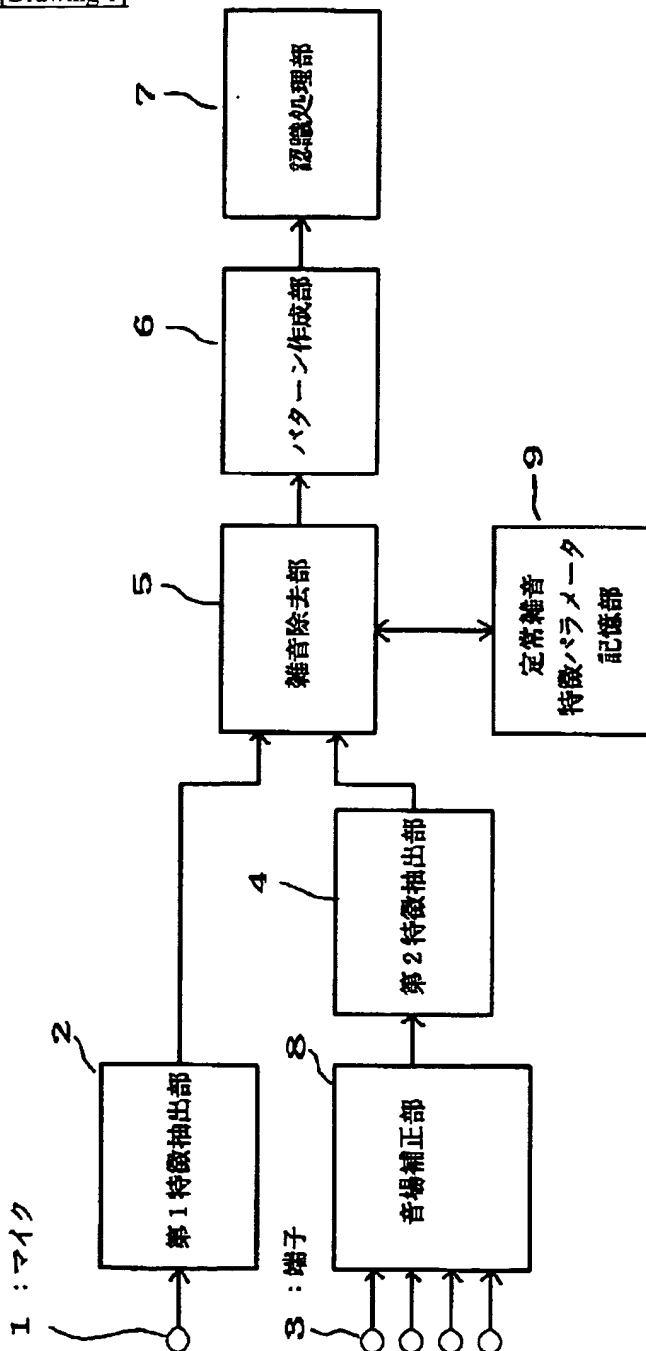

[Translation done.]

DRAWINGS

[Drawing 1]

7 認識処理部
6 パターン作成部
5 雑音除去部
9 定常雑音特徴パラメータ記憶部
4 第2特徴抽出部
2 第1特徴抽出部
8 音場補正部
1 : マイク
3 : 端子

[Drawing 2]
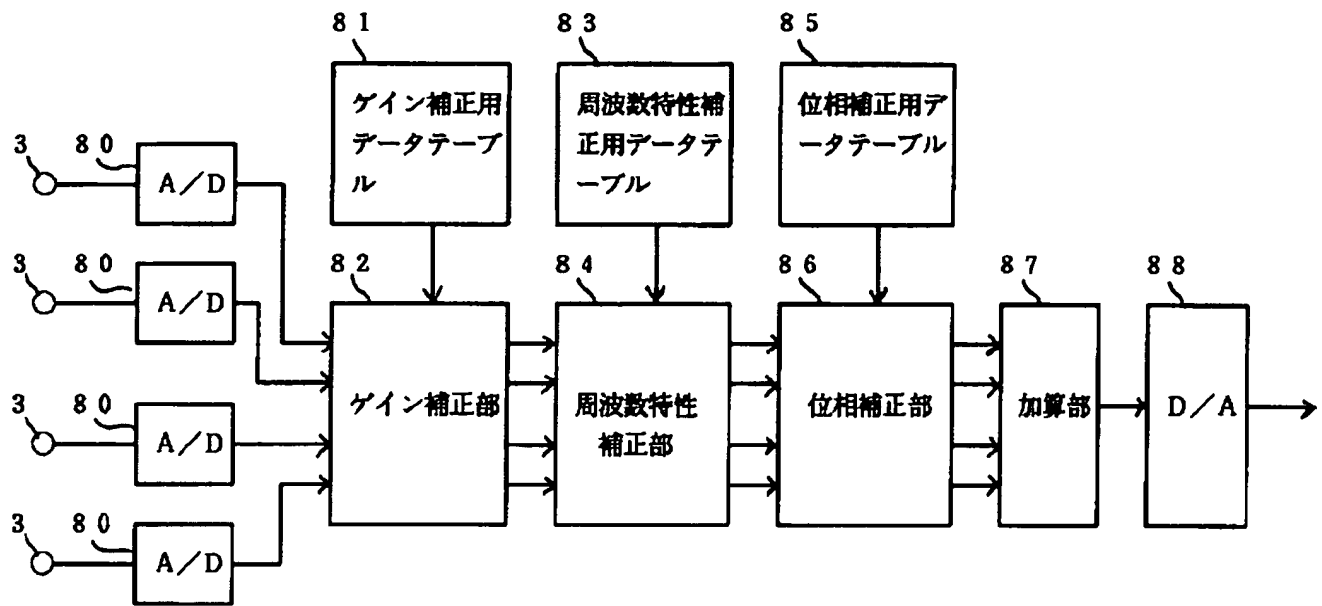
[Translation done.]

DESCRIPTION OF DRAWINGS

[Brief Description of the Drawings]
[Drawing 1] The basic block diagram of the speech recognition processor of this invention.
[Drawing 2] The block block diagram of the sound field amendment means 8 of this invention equipment.
[Description of Notations]
1 Microphone
2 1st Feature-Extraction Section
3 Noise Input Terminal
4 2nd Feature-Extraction Section
5 Normal-Mode-Rejection Section
6 Pattern Creation Section
7 Recognition Processing Section
8 Sound Field Amendment Section
82 Gain Amendment Section
84 Frequency Characteristic Amendment Section
86 Phase Correction Section

[Translation done.]

DETAILED DESCRIPTION

[Detailed Description of the Invention]
[0001]
[Industrial Application] Especially this invention relates to the voice recognition unit which eliminated the influence of the sound (it becomes noise for speech recognition) produced with sound equipments, such as television and radio, etc. about a voice recognition unit.
[0002]
[Description of the Prior Art] Although a speaker will utter voice toward a microphone and this voice will be first inputted into a microphone when performing speech recognition, as for a speaker's phonation environment, it is common at this time that do not necessarily restrict that it is good but a certain noise exists rather. Therefore, since noise is usually overlapped in the case of voice input and it is inputted in many cases at it, eliminating the influence of noise appropriately leads to realizing speech recognition with a high recognition rate.
[0003] Therefore, the various proposals of the method of removing such noise in recent years are made. In such a conventional normal-mode-rejection method, like a voice recognition unit given in for example, the patent application public presentation Showa No. 91007 [54 to] official report Using the 1st microphone for voice input, and the 2nd microphone for a noise input, from noise superposition sound signal N+V (sound signal V+ noise signal N) inputted from the 1st microphone, the noise signal N inputted from the 2nd microphone is deducted, and there is ** which is going to obtain the sound signal V without noise.
[0004] in such conventional technology, a setup of the installation position of the microphone for a noise input is difficult, for example, when the microphone for voice input near-boils the microphone for a noise input and it is installed, there is un-arranging [as which voice will be inputted into the microphone for a noise input comparatively greatly] Moreover, when installing the microphone for voice input, and the microphone for a noise input in the considerably distant position conversely, the properties (a phase, frequency, volume, etc.) of the noise inputted with the voice inputted into the microphone for voice input and the noise inputted into the microphone for a noise input were not in agreement, and even if it took these differences, there was un-arranging [not resulting] in an effective normal mode rejection.
[0005] It ** and these sound serves as noise for speech recognition depending on the environment where speech recognition is performed, in the environment where sound which audio equipments, such as television and radio, emit from the loudspeaker, such as music and voice, exists, for example. Thus, by coming out, then, since direct detection of television with a clear noise source or the radio acoustic signal which they emit can be carried out from the output terminal of these audio equipments, the above microphones for a noise input are omissible.
[0006] However, even when obtaining directly the acoustic signals N which such an audio equipment emits, such as music and voice, from these devices, there is often a case where the indirect sound voice Nz inputted into this direct acoustic signal N and the microphone for voice input for speech recognition with voice differs in property. For example, in in the car [of an automobile], since the music (acoustic noise) emitted from the loudspeaker of a car stereo is influenced of the sound property of the space of the automobile interior of a room, it has a fear of becoming the signal Nz which caused change to signal properties, such as a phase, frequency, and volume, and becoming a different thing in property from the signal N of the music (acoustic noise) obtained from the output terminal of this car stereo.
[0007] therefore, when signal properties with the indirect acoustic signal Nz which detected with the microphone the sound which was emitted from the signal property and loudspeaker of the direct acoustic signal N from the output terminal of an audio equipment, and was distorted by space environmental influence like **** differ Even if it deducts the direct acoustic signal N from the output terminal of an audio equipment from composite signal Nz+V of the indirect acoustic signal Nz and sound signal V which were inputted into the microphone, this difference signal V+Nz-N Since it did not become the true sound signal V, in such normal-mode-rejection processing, speech recognition was unrealizable in the quantity of a recognition rate after all.
[0008]
[Problem(s) to be Solved by the Invention] When sound properties with the indirect acoustic signal Nz which this invention is made in view of an above-mentioned point, serves as sound which was emitted from the signal property and loudspeaker of the direct acoustic signal N from the output terminal of an audio equipment, and was distorted by space environmental influence, and is detected with a microphone differ By considering the difference among these sound properties (a phase, frequency, volume, etc.) as an amendment (N=Nz), the true sound signal V is obtained and a voice recognition unit with a

high precision is realized.

[0009]

[Means for Solving the Problem] The microphone into which the above-mentioned acoustic noise is indirectly inputted through the above-mentioned space while the voice by which the voice recognition unit of this invention was uttered for speech recognition is inputted, The acoustic signal input terminal into which the signal of the above-mentioned acoustic noise is directly inputted from sound equipment, without minding the above-mentioned space, So that the property difference between the indirect signal of the acoustic noise obtained from the above-mentioned microphone and the direct signal of the acoustic noise obtained from the above-mentioned acoustic signal input terminal may be reduced The direct signal of an acoustic noise inputted from the above-mentioned acoustic signal input terminal A property amendment sound property amendment means, A 1st feature-extraction means to analyze the input signal from the above-mentioned microphone, and to extract the 1st feature parameter, A 2nd feature-extraction means to analyze the input signal rectified from the above-mentioned sound property amendment means, and to extract the 2nd feature parameter, A normal-mode-rejection means to deduct the 2nd feature parameter of the above from the 1st feature parameter of the above, and to generate the 3rd feature parameter, It has a voice pattern creation means to create a voice pattern from the 3rd feature parameter of the above, and a recognition processing means to perform recognition processing to the voice pattern created with the above-mentioned voice pattern creation means.

[0010]

[Function] By the above-mentioned sound property amendment means, the voice recognition unit of this invention amends the signal property of the direct acoustic signal N from the output terminal of an audio equipment so that it may be in agreement with the sound property of the indirect acoustic signal Nz detected with a microphone. That is, the characteristic distortion by space environmental influence is given to the direct acoustic signal N from the output terminal of an audio equipment, and it is equally to the indirect acoustic signal Nz, and is made properties (a phase, frequency, volume, etc.). Speech recognition can be performed based on the true sound signal V which deducted the signal property of the amendment direct acoustic signal Nz which rectifies the direct acoustic signal N of the output terminal of an audio equipment, and is obtained from the signal property of composite signal Nz+V of the indirect acoustic signal Nz and sound signal V which are detected with a microphone by this.

[0011]

[Example] The example at the time of using the speech recognition processor by this invention for below in the interior of a room of the automobile equipped with sound equipment is explained.

[0012] The example of 1 composition of the speech recognition processor by this invention is shown in drawing 1 . In drawing 1 , the voice as which the microphone into which 1 inputs voice was inputted into 2 is analyzed, and the 1st feature-extraction section which extracts the 1st feature parameter is shown. 3 is a noise input terminal which carries out the direct input of the output (acoustic noise) from the output terminal of sound equipment etc., and when four loudspeakers exist in sound equipment, four noise input terminals exist.

[0013] 4 is the 2nd feature-extraction section, after rectifying in the sound field amendment section 8 later mentioned to the direct signal acquired from the above-mentioned noise input terminal 3, analyzes noise and extracts the 2nd feature parameter.

[0014] In addition, as the above 1st and the 2nd feature-extraction section 2 and 4, the same analysis processing, for example, an FFT processor etc., is used.

[0015] 5 is the normal-mode-rejection section which deducts the 2nd feature parameter from the 1st feature parameter, and extracts the 3rd feature parameter.

[0016] 6 shows the pattern creation section which creates a voice pattern from the 3rd feature parameter obtained from the normal-mode-rejection section 5. Two or more voice patterns which are the recognition processing section which performs recognition processing of the voice pattern obtained from the pattern creation section 6, and serve as a candidate for recognition are already registered, and 7 performs audio specification, i.e., recognition, by finding out the voice pattern registered [analogous] most as a result of comparison with a strange voice pattern.

[0017] In addition, if needed, 9 is the stationary-noise feature-parameter storage section which memorizes the feature parameter of a stationary noise (the automobile interior of a room run sound, such as engine sound), and is used for normal-mode-rejection processing in the above-mentioned normal-mode-rejection section 5. That is, in case the 2nd feature parameter is deducted from the 1st feature parameter and the 3rd feature parameter is extracted, further, it puts, and the stationary-noise feature parameter of this storage section 9 will also be, will be lengthened, and the 3rd feature parameter will be obtained.

[0018] Here, the sound field amendment section by which the voice recognition unit of this invention is characterized most is explained in full detail. This sound field amendment section 8 is a property amendment sound property amendment means about the direct signal of an acoustic noise inputted from the above-mentioned noise input terminal 3, as the difference of the properties between the indirect signal of the acoustic noise obtained from the above-mentioned microphone 1 and the direct signal of the acoustic noise obtained from the above-mentioned noise input terminal 3 (a phase, frequency, volume, etc.) is reduced.

[0019] The composition of the sound field amendment section 8 as such a sound property amendment means is shown in drawing 2 . in this drawing -- 80 and 80 -- it is the A/D converter which changes into a digital signal the acoustic-noise signal into which .. was inputted from an analog signal

[0020] The data table for a gain amendment to which 81 holds the data for a gain amendment, the data table for a frequency characteristic amendment to which the amendment gain amendment section and 83 hold the data for a frequency characteristic amendment for gain using this data for a gain amendment in 82, the data table for phase corrections to which the amendment frequency characteristic amendment section and 85 hold the data for phase corrections for the frequency characteristic using this data for a frequency characteristic amendment in 84, and 86 are the amendment phase correction sections about a phase using the data for phase corrections.

[0021] Furthermore, 87 is an adder unit adding each four outputs from the phase correction section 86 to all four loudspeakers, and 88 is a D/A converter which changes the digital signal of this aggregate value into an analog signal.

[0022] In such the sound field amendment section 8 of composition, it precedes performing speech recognition, the difference in the property of the sound data produced between the microphone 1 for voice and the noise input terminal 3 is measured, and creation of the data for a property amendment is performed.

[0023] The pulse sound which sound equipment was made to generate pulse sound, was specifically outputted to the microphone 1 from the loudspeaker, and received space environmental influence of an automobile, such as the interior of a room, is inputted, and a direct input is carried out from the output terminal of sound equipment in the form of an electrical signal to the noise input terminal 3. At this time, in the gain amendment section 82, adjust volume, adjust the frequency characteristic in the frequency characteristic amendment section 84, adjust a phase in the phase correction section 86, and it sets in the normal-mode-rejection section 5 as a result of these the adjustments of each. Each value of the data table 81 for a gain amendment, the data table 83 for a frequency characteristic amendment, and the data table 85 for phase corrections is set up so that the difference (the 3rd feature parameter) of a microphone 1 and the feature parameter obtained from the both sides of the noise input terminal 3 may serve as the minimum.

[0024] That is, in the gain amendment section 82, the frequency characteristic amendment section 84, and the frequency characteristic amendment section 84, when amendment regulation of each parameter is performed, and the 3rd feature parameter does not exist or it becomes the minimum, feeding back the difference segregation malleolus of the normal-mode-rejection section 5, the amendment data (a constant or suitable function) of each parameter are memorized to the data table 81 for a gain amendment, the data table 83 for a frequency characteristic amendment, and the data table 85 for phase corrections.

[0025] According to drawing 1 and drawing 2 , explanation of recognition processing operation is added to below.

[0026] First, the audio voice analysis inputted from the microphone 1 is performed by the 1st feature-extraction section 2. In the 1st feature-extraction section 2 of KO, after changing the inputted analog sound signal into a digital signal, the 1st feature parameter is extracted by performing the Fourier transform to the obtained digital signal. The 1st feature parameter obtained here is further sent to the normal-mode-rejection section 5.

[0027] Moreover, while voice is inputted from a microphone 1, the acoustic noise of a car stereo is directly inputted into the noise input section 3 from the output terminal. For example, when four loudspeakers exist, the input of the acoustic noise to the noise input terminal 3 is performed through four input terminals.

[0028] Four acoustic-noise signals inputted into these four input terminals 3 are sent to the sound field amendment section 8. The sound field amendment section 8 will rectify the difference between the phase produced between the acoustic-noise signals inputted from an acoustic-noise signal and a microphone 1, the frequency characteristic, and volume to each these four line.

[0029] That is, in the sound field amendment section 8, as shown in drawing 2 , A/D converter 80 changes each inputted noise signal into digital value first.

[0030] And the gain amendment section 82 rectifies gain to four outputs each from A/D converter 80 using the data for a gain amendment currently recorded on the data table 81 for a gain amendment.

[0031] Next, the frequency characteristic amendment section 84 rectifies the frequency characteristic to four outputs each from the gain amendment section 82 using the data for a frequency characteristic amendment memorized by the data table 83 for a frequency characteristic amendment.

[0032] Then, the phase correction section 86 rectifies a phase to four outputs each from the frequency characteristic amendment section 84 using the data for phase corrections currently recorded on the data table 85 for phase corrections.

[0033] After these amendments are completed, each four outputs from the phase correction section 86 of the last stage are above-mentioned [an adder unit 87] in the stage are added to one acoustic-noise signal, and D/A converter 88 changes this into an analog signal.

[0034] Above, processing of the sound field amendment section 8 is completed. In the 2nd feature-extraction section 4, the acoustic-noise signal rectified by the sound field amendment section 8 as mentioned above performs the same voice-analysis processing as the above-mentioned 1st feature-extraction section 2, and extracts the 2nd feature parameter.

[0035] The 1st feature parameter created from the input of a microphone 1 and the 2nd feature parameter created from the input of the noise input section 3 are transmitted to the normal-mode-rejection section 5.

[0036] The normal-mode-rejection section 5 deducts fundamentally the 2nd feature parameter which consists of an acoustic noise from the 1st feature parameter which consists of an acoustic noise and voice, and generates the 3rd feature parameter which does not contain an acoustic noise.

[0037] Furthermore, the pattern creation section 6 detects the voice section of the 3rd feature parameter, and creates a voice pattern.

[0038] The recognition processing section 7 performs recognition processing by comparing the voice pattern created by doing

in this way with the voice pattern registered into the recognition processing section 7.

[0039] Thus, the fault which according to this invention originates in the position of a noise source, the position of the microphone for voice input, etc., and is produced, namely, the property (a phase --) produced between the feature parameters of noise inputted from the microphone after receiving this space environmental influence, while the feature parameter of the actual noise itself and actual noise had spread the inside of space differences, such as frequency and volume, -- an amendment -- things become possible, noise is certainly removed in the environment where acoustic noises, such as in the car [of the automobile especially equipped with the car stereo], always change, and the recognition performance degradation by noise is prevented

[0040] In addition, also in other space, although this example described the case where the voice recognition unit by this invention was used in the in-the-car space of the automobile equipped with the car stereo, before inputting the sound data of noise into a voice recognition unit through speech recognition space, in any environment which serves as known by another means, it is applicable [this invention].

[0041] Moreover, in order to raise recognition precision in the speech recognition processor which operates under environment which was mentioned above, it is required to also often take into consideration noises (stationary noise), such as engine sound, run sound, and sound made according to a natural phenomenon. Explanation is added about the case where such a stationary noise is removed in this speech recognition processor below.

[0042] Fundamentally, removal of a stationary noise is made from the 3rd feature parameter obtained in the normal-mode-rejection section 5 according to the change by deducting the feature parameter of the stationary noise of the stationary-noise parameter storage section 9. Here, a stationary-noise parameter points out the thing of the 1st feature parameter created in the state where the input of the acoustic noise to the noise input terminal 3 does not exist, and the voice to a microphone 1 and the input of an acoustic noise do not exist.

[0043] What is necessary is to be in the state which is not outputted at all from sound equipment, to extract the 1st feature parameter and just to specifically, memorize this in the stationary-noise parameter storage section 9 beforehand, by this example, since sound equipment is used as a noise source.

[0044] Therefore, at the time of recognition, the input from the noise input terminal 3 is processed from a microphone 1 to an input list by method which was already described, and the 1st feature parameter and the 2nd feature parameter are extracted.

[0045] The normal-mode-rejection section 5 deducts the 2nd feature parameter from the 1st feature parameter, further, deducts the stationary-noise parameter of the stationary-noise parameter storage section 9, and creates the 3rd feature parameter. Like ****, the pattern creation section 6 detects the voice section to the 3rd feature parameter with which the stationary-noise parameter is deducted, and creates a voice pattern. The recognition processing section 7 performs recognition processing by comparing the voice pattern created by doing in this way with the voice pattern registered into the recognition processing section 7.

[0046] Moreover, since the stationary noise made by the engine noise etc. may change gently, in order to remove a stationary noise effectively in this case, it is desirable to update the stationary-noise parameter of the stationary-noise parameter storage section 9 if needed. Such an update process is described below.

[0047] if the value of introduction and the 3rd feature parameter is less than [this] first, the circumference noise or voice from a car stereo are not inputted into a speech recognition processor -- ** (an input is only a stationary noise) -- the threshold which can be concluded is set up

[0048] Therefore, the normal-mode-rejection section 5 checks the value of the 3rd feature parameter at every fixed period (for example, 10msec(s)) at the time of recognition, when the value of the 3rd parameter is below a threshold, considers that the stationary noise was inputted from the microphone, and updates a stationary-noise parameter, without performing detection of voice input.

[0049] such an updating method is new in the result which carried out the weighted mean of for example, a stationary-noise parameter and the 3rd feature parameter by ratio to which the rate of a stationary-noise parameter becomes high -- a stationary-noise parameter is carried out and it registers

[0050] In fact, the ratio of a stationary-noise parameter is enlarged and carries out a weighted mean so that a big change may not arise in the value of a stationary-noise parameter (by for example, 15:1 ratios).

[0051] Thus, since it can be effectively coped with also by steady noises, such as an engine noise in an automobile, according to this invention, the recognition performance degradation by noise is prevented further.

[0052] In addition, in this example, although the frequency characteristic was changed by FFT and IFFT, it is also possible to carry out by other meanses. It is also possible to perform a sound field amendment after a feature extraction. Moreover, in this example, although the direct input of the input to the noise input section was carried out from the loudspeaker, it is also possible to install and process a microphone so that desired voice may not be inputted. The composition by the circuit and realization by DSP are also possible for these processings.

[0053]

[Effect of the Invention] Since the voice recognition unit of this invention gives the characteristic distortion by space environmental influence to the direct acoustic signal N from the output terminal of an audio equipment, and is equally to the indirect acoustic signal Nz of a microphone input and it is made to properties (a phase, frequency, volume, etc.) From the signal property of composite signal Nz+V of the indirect acoustic signal Nz and sound signal V which are detected with a microphone Based on the true sound signal V which deducted the signal property of the amendment direct acoustic signal Nz which rectifies the direct acoustic signal N of the output terminal of an audio equipment, and is obtained, speech recognition

can be performed and speech recognition with a high recognition rate can be realized also under noise environment.

[Translation done.]